Модель обслуживания одиночной заявки в сети с узлами, нагруженными другими внешними источниками

Single application service model in a network with nodes loaded by other external sources

Смагин / Smagin V.
Владимир Александрович
(va_smagin@mail.ru)
доктор технических наук, профессор,
заслуженный деятель науки РФ,
действительный член МАИ.
ФГБВОУ ВО «Военно-космическая академия
имени А. Ф. Можайского» МО РФ,
профессор кафедры метрологического обеспечения.
г. Санкт-Петербург

Ключевые слова: сеть массового обслуживания – mass service network; очереди в узлах – lines in nodes; заявка одиночного потребителя – single consumer's application; матричный метод – matrix method; произвольные распределения вероятностей – generalized distribution of probabilities; преобразование Лапласа – Laplace transform; формула Алфрея – Alfrey's formula.

Предложена математическая модель обслуживания заявки в вычислительной сети, поступающей от определённого заинтересованного потребителя. В отличие от известных моделей сетей, в которых заявки, поступающие на обслуживание в любой узел, начинают обслуживаться мгновенно, в данной сети указанная заявка в любом узле вынуждена пройти время ожидания. Времена ожиданий в узлах обусловлены тем, что все узлы сети обслуживают других внешних источников.

A mathematic model is proposed for serving application in computation network being delivered by a specific consumer interested. As against familiar network models, in which applications being delivered for service in any node are commenced to be served immediately, in present network said application in any node is forced to pass waiting time. Waiting times in the nodes are conditioned by the fact that any network node is served from other external sources.

Введение

Современные сложные сетевые системы содержат множество узлов, выполняющих различные функции. Временные процессы, происходящие в них, зависят от ряда внешних и внутренних факторов. Выбор наиболее важных из этих факторов определяет условия и особенности моделирования процесса функционирования сетей. Моделирование работы сложных сетей наиболее продуктивно на средствах вычислительной техники. Разработка новых моделей функционирования сетей, отличающихся от известных более сложной динамикой происходящих в них процессов, требует от исследователя постановки и решения новых математических задач и алгоритмов. Поэтому принятие во внимание новых физических аспектов в работе реальных сетей, развитие известных и построение новых аналитических моделей, отражающих влияние ранее не учитываемых факторов, является актуальной задачей исследователя.

В ряде работ, посвящённых обслуживанию поступающих требований, заявок в сетях, как правило, не учитывается динамизм работы узлов, обусловленный их загруженностью заявками другого характера. Поэтому новая одиночная заявка, учитывающая интерес конкретного потребителя и приходящая в сеть для обслуживания именно для его целевого удовлетворения [1], вынуждена пробивать себе брешь в мешающем, не нужном ему динамизме работы узлов сети. В качестве элементарного примера представим себе покупателя, приходящего в магазин для совершения покупки определённого ассортимента товаров. Он должен пройти на своём пути ряд узлов обслуживания товаров, уже занятых очередями покупателей.

Описание процесса обслуживания в сети. Одиночная заявка поступает в узел-исток сети, содержащей некоторое конечное множество узлов обслуживания. Любой узел сети, кроме узлов истока и стока, находится в состоянии непрерывного динамического обслуживания заявок, поступающих от других источников.

Он может иметь на своём входе случайное количество заявок, ожидающих обслуживания. Одиночная заявка со входа сети, пришедшая в любой узел динамического обслуживания поступивших уже заявок, становится в очередь и находится в ней в течение времени ожидания своего обслуживания. Переходы между узлами задаются матрицей вероятностей переходов. Процесс обслуживания в динамической сети заканчивается, когда одиночная заявка достигает узла-стока. Интерес представляет нахождение вероятностно-временных характеристик обслуживания одиночной заявки в данной динамической сети.

Цель настоящей статьи построить при некоторых дополнительных функциональных особенностях работы отдельных узлов сети, математическую модель, позволяющую определять приближённо вероятностные характеристики обслуживания заявки, поступающей в сеть с более сложной, динамической архитектурой.

О предшествующих исследованиях. В статье [2] рассматривался приближённый метод решения уравнения Винера-Хопфа для произвольных гладких распределений времени между поступающими на вход узла заявками на обслуживание и времени их обслуживания. Метод основан на гипердельтной аппроксимации распределений [3], которая скорректирована с целью её применения для случайных величин, сосредоточенных на всей вещественной оси абсцисс [4]. Для коррекции применён метод Фурье и характеристические функции. Их использование позволило приближённо, путём последовательных итераций находить функцию распределения времени ожидания обслуживания поступающей заявки в узел. Однако в сложной системе, состоящей из нескольких узлов, связанных между собой вероятностными связями, вопрос определения времени ожидания для всей системы остаётся нерешённым.

Математическая модель сети. Рассмотрим, как и в [5], сеть обслуживания (СО), состоящую из M-узлов. Она представляется в виде ориентированного графа, нулевая вершина которого – исток – идентифицирует источник одиночной заявки, вершины 1,2,..., M – узлы сети, а (M+1)-я вершина – выход, или – сток сети. Переходы между узлами определены матрицей вероятностей передач сети r_{ij} со стационарным режимом работы. Плотности вероятностей времени пребывания заявок в очередях узлов задаются в форме, удовлетворяющей производной по времени от функции распределения времени ожидания. Они должны определяться в результате решения интегрального уравнения Винера-Хопфа. Однако в замкнутом аналитическом виде для произвольных законов распределений решение уравнения не получено. Поэтому распределение времени ожидания можно принимать лишь в некоторых известных случаях. К таким случаям относятся полученные в замкнутом виде решения для времени ожидания уравнения Винера-Хопфа и формула Полячека-Хинчина. Например, когда оба исходные распределения являются экспоненциальными, либо одно экспоненциальное, а другое таковым не является. Наибольшее внимание привлекает распределение времени ожидания в форме Полячека-Хинчина, при котором входящий поток пуассоновский, а распределение обслуживания – произвольное. Достаточно сложный вид имеет распределение времени ожидания при входящем пуассоновском потоке и постоянном времени обслуживания и некоторых других частных случаях. В нашем исследовании обслуживания в динамической (СО) мы за исходную форму принимаем изображение Лапласа функции времени ожидания Полячека-Хинчина [6]:

$$\omega(s) = \frac{1-\rho}{1-(\lambda/s)[1-\beta(s)]}, \tag{1}$$

где $\omega(s), \beta(s)$ – изображения Лапласа функции распределения времени ожидания в очереди и плотности вероятности времени обслуживания заявки, s – переменная Лапласа, λ – интенсивность пуассоновского потока, ρ – коэффициент загрузки узла.

Изображение Лапласа плотности вероятности времени пребывания заявки в i-ом узле сети с учётом ожидания начала обслуживания будет равно:

$$g_i(s) = \omega_i(s)\beta_i(s) = \frac{(1-\rho_i)\beta_i(s)}{1-(\lambda_i/s)[1-\beta_i(s)]}. \tag{2}$$

Следует отметить, что плотность вероятности, определяемая как оригинал обратного преобразования Лапласа, является разрывной, что необходимо принимать во внимание при производстве аналитических выкладок в расчётах характеристик сети. Аппроксимационный метод расчёта закона распределения времени пребывания заявки в узлах произвольной разомкнутой СО определён в [7]. Предполагалось, что СО удовлетворяет условиям локального баланса и её решение может быть представлено в форме произведения [8]. Задача состояла в определении закона распределения времени пребывания заявки в сети или между i-м и j-м узлами данной сети, $i = 0,1,\ldots,M+1;\ j = 0,1,\ldots,M+1;\ i \neq j$.

Метод решения задачи основан на предположении о вероятностной независимости времени пребывания заявки в j-ом узле от времени пребывания в узле i. Это предположение основывается на локальности баланса и стационарности режима функционирования сети.

Введём в рассмотрение матрицу

$$G(s) = \begin{bmatrix} r_{0,0} & r_{0,1} & \cdots & r_{0,M} & r_{0,M+1} \\ r_{1,0}g_1(s) & r_{1,1}g_1(s) & \cdots & r_{1,M}g_1(s) & r_{1,M+1}g_1(s) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{M,0}g_M(s) & r_{M,1}g_M(s) & \cdots & r_{M,M}g_1(s) & r_{M,M+1}g_M(s) \\ r_{M+1,0} & r_{M+1,1} & \cdots & r_{M+1,M} & r_{M+1,M+1} \end{bmatrix}. \tag{3}$$

Каждый элемент матрицы $r_{ij}g_i(s)$ есть произведение изображения Лапласа плотности распределения времени пребывания заявки в i-ом узле сети на вероятность перехода после обслуживания в i-ом узле в узел j. Время пребывания в 0 и M+1 узлах равно нулю, поэтому $g_0(s) = g_{M+1}(s) = 1$.

Изображение Лапласа плотности распределения времени пребывания заявки между i и j узлами сети равно элементу (i, j) матрицы $T(s)$ вида

$$T(s) = I + G(s) + G^2(s) + \cdots = I(I - G(s))^{-1}, \qquad (4)$$

где I – единичная матрица.

Элемент матрицы $T(s)$ с номером $(0, M+1)$ представляет собой изображение Лапласа плотности вероятности времени пребывания в сети с учётом всех возможных траекторий следования к её стоку.

В соответствии с правилом вычисления обратной матрицы элемент (i, j) равен

$$Y_{i,j}(s) = A_{j,i}(s)/R(s), \qquad (5)$$

где $A_{i,j}(s)$ – алгебраическое дополнение элемента (i, j) матрицы $T(s)$, $R(s)$ – определитель матрицы $T(s)$. Из (5) следует, что изображение Лапласа времени нахождения заявки в сети соответственно равно

$$Y_{0,M+1}(s) = A_{M+1,0}(s)/R(s). \qquad (6)$$

Если время пребывания заявки в каждом узле сети задать начальными моментами (порядок их определения при самых общих предположениях о входящем потоке и распределении времени обслуживания приведён в [7]), то можно предложить алгоритм определения начальных моментов времени пребывания заявки между i-м и j-м узлами сети. Алгоритм основан на связи между k-м начальным моментом и k-й производной изображения Лапласа в точке $s = 0$. Подставляя вместо $\left.\frac{d^k g_l(s)}{ds^k}\right|_{s=0}$ $l = 0,1,\ldots,M+1$, значение k-го момента распределения времени нахождения заявки в l-м узле сети, а вместо $\left.g_l(s)\right|_{s=0}$ единицу, получаем k-й начальный момент распределения времени нахождения заявки между i-м и j-м узлами сети. По вычисленным начальным моментам можно известными методами получить аппроксимирующую плотность распределения времени пребывания заявки между i-м и j-м узлами СО.

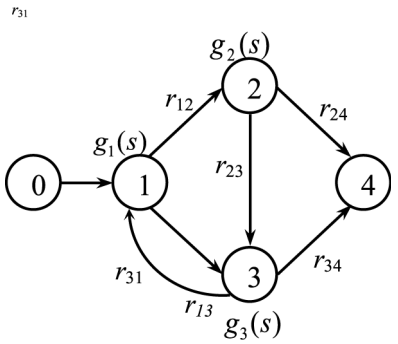

Рис. 1

Пример простой структуры сети. На рисунке 1 представлена структура сети, включающая три узла обслуживания с задержками одиночной заявки. Исходные данные для расчёта: переходные вероятности между узлами $r_{12} = 0{,}7$; $r_{13} = 0{,}3$; $r_{31} = 0{,}2$; $r_{23} = 0{,}4$; $r_{24} = 0{,}6$; $r_{34} = 0{,}8$. Изображения Лапласа плотностей вероятностей времен пребывания заявок в узлах сети с учётом задержек на время ожиданий в очередях: в узле 1 – простейший входящий поток, обслуживание – двухфазная гамма плотность, время ожидания обслуживания распределено по закону Полячека-Хинчина. Поэтому

$$g_1(s) = \omega_1(s)\beta_1(s) = \frac{(1-\rho_1)\beta_1(s)}{1-(\lambda_1/s)[1-\beta_1(s)]},$$

$$\lambda_1 = 0{,}02\,h^{-1},\ \beta_1(s) = \frac{\mu_1^2}{(s+\mu_1)^2},\ \mu_1 = 0{,}1\,h^{-1},\ \rho_1 = \frac{2\lambda_1}{\mu_1}. \qquad (7)$$

В узле 2 – простейший входящий поток, экспоненциальное распределение времени обслуживания. Изображение Лапласа плотности вероятности времени пребывания в узле

$$g_2(s) = \frac{(1-\rho_2)(s+\mu_2)\mu_2}{s(s+\mu_2-\lambda_2)(s+\mu_2)} = \frac{(1-\rho_2)\mu_2}{s(s+\mu_2-\lambda_2)},$$

$$\rho_2 = \frac{\lambda_2}{\mu_2},\ \lambda_2 = 0{,}05\,h^{-1},\ \mu_2 = 0{,}15\,h^{-1}. \qquad (8)$$

Узел 3 имеет пуассоновский входящий поток, постоянное время обслуживания, поэтому изображение Лапласа плотности вероятности времени пребывания в узле принимает вид

$$g_3(s) = \frac{(1-\rho_3)\beta_3(s)}{1-\frac{\lambda_3}{s}(1-e^{-sb^{-1}})},\ \beta_3(s) = e^{-sb^{-1}},$$

$$\rho_3 = \frac{\lambda_3}{b},\ \lambda_3 = 0{,}07\,h^{-1},\ b = 0{,}2\,h^{-1}. \qquad (9)$$

В рассматриваем случае $M = 3$. Составим матрицы $G(s)$ и $(I - G(s))$:

$$G(s) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & r_{12}g_1(s) & r_{13}g_1(s) & 0 \\ 0 & 0 & 0 & r_{23}g_2(s) & r_{24}g_2(s) \\ 0 & r_{31}g_1(s) & 0 & 0 & r_{34}g_3(s) \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \qquad (10)$$

$$(I-G(s))=\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -r_{12}g_1(s) & -r_{13}g_1(s) & 0 \\ 0 & 0 & 1 & -r_{23}g_2(s) & -r_{24}g_2(s) \\ 0 & -r_{31}g_3(s) & 0 & 1 & -r_{34}g_3(s) \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (11)$$

Элемент матрицы $(I-G(s))^{-1}$ с номером согласно (5) будет равен $Y(s)=Q(s)/R(s)$, где $Q(s)$ – алгебраическое дополнение элемента матрицы $(I-G(s))$. Раскрывая данные определители, получим следующее выражение

$$Y(s)=\frac{r_{12}g_1(s)r_{23}g_2(s)r_{34}g_3(s)+}{1-r_{13}g_1(s)r_{31}g_3(s)-}$$

$$\frac{+r_{12}g_1(s)r_{24}g_2(s)+r_{13}g_1(s)r_{34}g_3(s)}{-r_{12}g_1(s)r_{23}g_2(s)r_{31}g_3(s)}. \quad (12)$$

Ранее при выполнении исследования сетей без принятия во внимание времени ожидания обслуживания поступившей заявки, когда она начинала обслуживаться сразу после поступления на вход узла, выполнялось дифференцирование (12) по переменной s и при s=0 находились начальные моменты случайного времени пребывания в сети. Затем по найденным начальным моментам находилось искомое распределение вероятностей.

В данном случае такой подход неприменим. Это объясняется тем, что аналитическое дифференцирование (12) по переменной s выполнить практически не представляется возможным. Поэтому следует получать искомый результат, представленный во временной области, другим способом, который основан на приближённом обращении преобразования Лапласа. Для этого рекомендуется в [9] применить формулы Хаара или Алфрея. Эти формулы следуют из формулы Уайдера – формулы приближённого обращения изображения Лапласа на основе свойства фильтрации преобразования с помощью дельта-функции.

Замечание. Выполним, например, приближённое обращение (12), пользуясь аппроксимационной формулой Алфрея, которая принимает следующий вид:

$$f(t)\approx sf^*(s) \text{ при } s=\frac{1}{t}. \quad (13)$$

Для пояснения приведём простейший пример из теории надёжности. Изображение вероятности безотказной работы при экспоненциальном законе $P^*(s)=\frac{1}{s+\lambda}$. По формуле Алфрея имеем $P(t)\approx sP^*(s)\Big|_{s=\frac{1}{t}}=\frac{1}{1+\lambda t}$. Это простейшая начальная аппроксимации вероятности. Можно убедиться, что в данном случае она даёт завышенный результат и тем больше, чем больше параметр закона распределения λ. Представление вероятности будет тем точнее, чем меньше величина времени t.

Далее, вернёмся к нашей задача и найдём выражение $y(t)=s\,Y(s)\Big|_{s=\frac{1}{t}}$.

Вспомним, что по формуле (12) мы вычисляем изображение Лапласа плотности вероятности времени пребывания одиночной заявки, поступившей на вход сети, для успешного её обслуживания структурой сети с временами ожиданий обслуживания в узлах. Поэтому изображение Лапласа вероятности неуспешного обслуживания будет определяться по известной формуле, подобной выше приведённой формуле для экспоненциального закона

$$P^*(s)=\frac{1-Y\ (s)}{s} \quad (14)$$

Тогда оригинал вероятности выполнения нашей задачи во временной области будет приближённо определён по формуле Алфрея

$$P(t)\approx s\,P^*(s)\Big|_{s=\frac{1}{t}}. \quad (15)$$

Выполнив все необходимые аналитические преобразования и проведя численные расчёты при указанных исходных данных, окончательно приведём график искомой вероятности в зависимости от времени, который представлен на рисунке 2.

Ещё раз обратим внимание, что на рисунке представлена вероятность неуспешного обслуживания заявки от времени сплошной кривой и успешного обслуживания – пунктирной кривой. Вероятность успешного обслуживания заявки будет представлена формулой

$$Q(t)=1-P(t). \quad (16)$$

Также приведём численные значения отдельных параметров результата решения задачи. Найдём среднее время решения задачи, второй начальный момент, среднеквадратическое отклонение, коэффициент вариации. Они будут равны:

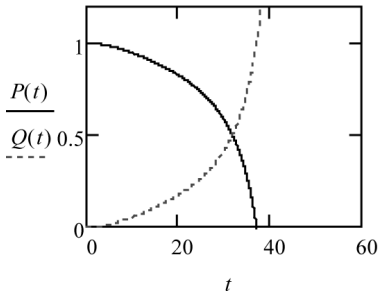

Рис. 2

$$T = \int_0^{38} P(z)dz = 28,51h;\ \alpha = \int_0^{38} za(z)dz = 1134,51h^2;$$

$$\sigma = 17,94h;\ \eta = 0,625.$$

По параметрам T, σ построим аппроксимирующую нормальную плотность вероятности, которая показана на рисунке 3. Она сосредоточена (-30,90)h.

Итак, достоверно, то есть с вероятностью единица, заявка получит полное обслуживание в рассматриваемой сети за 38 часов её функционирования. Несоответствие между поведением «хвостов» на рисунках 2 и 3, а именно резкая конечность на рисунке 2, определяется грубостью аппроксимации (линейной) оригинала функции формулой Алфрея. Требуется более точное выражение для перехода от изображения Лапласа к оригиналу во времени. Положительным свойством использованного перехода является чрезмерная простота преобразования для сложных аналитических выражений, позволяющая получать достаточно точные оценки показателей на ограниченных интервалах оценивания значений показателей.

Заключение

Рассмотрена более сложная задача для информационной сети по сравнению с ранее известными. Если ранее заявка, покидая узел обслуживания, мгновенно начинала обслуживаться в другом узле – адресате, то в данном случае она поступает в узел – адресат, находящийся уже в течение некоторого времени в состоянии обслуживания других заявок от некоторого (некоторых) другого источника заявок. В нашей терминологии одиночная заявка поступает в динамический узел.

Для решения задачи использован традиционный вероятностно-временной матричный аппарат. Однако вместо нахождения начальных моментов времени пребывания заявки в сети и аппроксимации искомого показателя методом моментов, использован метод приближённого перехода от изображений Лапласа к оригиналам на основе свойства фильтрации дельта-функцией прямого преобразования Лапласа.

Метод оказывается достаточно результативным в работе со сложными изображениями. Об этом свидетельствует рассмотренный в статье пример.

Однако для повышения точности вычислений в дальнейшем требуется искать более строгие методы обратного преобразования Лапласа на основе названных формул Хаара и Уайдера.

Но всё-таки наиболее важным вопросом для решения подобных сетевых задач этим методом остаётся представление решения уравнения Винера-Хопфа и формулы Полячека-Хинчина для произвольных, непуассоновских, входящих распределений.

Литература

1. Cheung, R. C. A User – Oriented Software Reliability Model / R.C. Cheung // IEEE Transactions On Software Engineering. – 1980. – Vol. 6. – ISS. 2. – P. 118–125.
2. Смагин, В. А. Решение интегрального уравнения Винера-Хопфа методом гипердельтной аппроксимации / В.А. Смагин // Intellectual Technologies on Transport. – 2016. – № 1. – С. 49–53.
3. Смагин, В. А. О моделировании случайных процессов на основе гипердельтного распределения / В.А. Смагин, Г.В. Филимонихин // АВТ. – 1990. – № 1. – С. 25–31.
4. Смагин, В. А. Коррекция гипердельтного распределения в теории случайных процессов / В.А. Смагин // Информация и Космос. – 2015. – № 4. – С. 60–64.
5. Смагин, В. А. Расчёт вероятностно-временных характеристик пребывания заявок в сетевой модели массового обслуживания / В.А. Смагин, В.П. Бубнов, Г.В. Филимонихин // Изв. Вузов. Приборостроение. – 1989. – № 2. – С. 23–25.
6. Риордан, Дж. Вероятностные системы обслуживания / Дж. Риордан . – М: Связь, 1966. – 184 с.
7. Смагин, В. А. Аппроксимационный метод расчёта разомкнутых сетей массового обслуживания / В.А. Смагин, Г.В. Филимонихин // АВТ. – 1986. – № 4. – С. 28–33.
8. Gelenbe, E. Analysis and Synthesis of computer systems / E. Gelenbe, I. Mitrani. – Academic Press, 1980. – 239 p.
9. Смагин, В. А. Немарковские задачи теории надёжности / В.А. Смагин. – МО СССР, 1982. – 269 с.

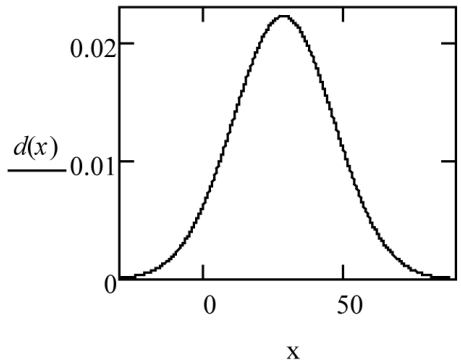

Рис. 3